山形大学紀要（工学）第22巻　第2号　平成5年1月
Bull. Yamagata Univ. (Eng.) Vol.22 No. 2 Jan. 1993

Unequivalent Conversion in Processing of Literature Data Incorporated in Systems for Estimation of Aqueous Solubility of Organic Compounds.

Yoshihiro KUDO

Department of Electrical and Information Engineering, Facluty of Engineering.

(*Received September* 1, 1992)

Abstract

Many experimental values have been used through citation as literature data. However during their processing, unequivalent conversion is frequently done. The present paper analysed our collection of articles treating aqueous solubility data and found several causes of this undesired event. Sometimes a value which was described as an original data in a paper does not exist in the paper where it was to be. Such unreasonable citation causes wrong phylogenesis of values. Clearly some of them were due to simple confusion of compound names. It is concluded that any data should be as faithfully transmitted as possible, and that necessary revision should be carried out in the manner whose histories are hold.

1. Introduction

We have developed some chemical inference systems, e.g., for aqueous solubility estimation[1], structure elucidation[2], and others. We have to always depend on literature values because of our inability to measure necessary values by ourselves. Inevitable variation among C13- and H-nmr data collected for our structure elucidation system were ascribed to natural measurement error, sensitivity of certain kinds of compounds, unconcious change of measurement condition, and human editorial mistake.[3] Another such situation is observed in construction of aqueous solubility databases containing literature values. Since aqueous solubility is very important from various viewpoints, many estimation systems have been developed on the basis of various kinds of physico-chemical theories. Nevertheless experimental values may increase in importance but would never decrease in. The present paper examines relationship of citation not only among individual values but also among articles treating them from a viewpoint of data reliability.

2. Relationship among articles that were citing and cited data

We have developed a system for aqueous solubility of organic compounds through estimation from their structure and/or a numerical database[1] of literature values. Table 1 is a part of our collection. A reference

Table 1. Collection of articles on aqueous solubility (part).
(An asterisked article is yet to be available.)

1) C. Hansch, J. E. Quinlan and G. L. Lawrence, J. Org. Chem., 3340, 347(1968).
*2) A. Seidell, "Solubilities of Organic Compounds," Vol.2, 3rd ed., D.Van Nostrand Co., New York, 1941.
3) P. W. Gross and J. H. Saylor, J. Am. Chem. Soc.,53.1744(1031).
4) P. W. Gross, J. H. Saylor and H. A. Gorman, J. Am. Chem. Soc., 55, 850(1933).
5) H. Fuhner, Ber., 57, 510(1924).
6) J. A. V. Butler, D. W. Thompson and W. H. Maclennan, J. Chem. Soc., 1933, 674.
7) P. W. Ginnings and R. Baum, J. Am. Chem. Soc., 59, 1111(1937).
8) P. W. Ginnings and R. Webb, J. Am. Chem. Soc., 60, 1388(1938).
9) P. W. Ginnings and W. Hauser, J. Am. Chem. Soc., 60 ,2581(1938).
10) P. W. Ginnings and D. Coltrane, J. Am. Chem. Soc., 61, 525(1939).
11) P. W. Ginnings, E. Herring and D. Coltrane, J. Am. Chem. Soc., 61, 807(1939).
12) P .W. Ginnings, D. Plunk and E. Carter, J. Am. Chem. Soc., 62, 1923(1940).
13) A. P. Altshuller and H. E. Everson, J. Am. Chem. Soc., 75, 1727(1953).
14) H. Sobotka and J. Kahn, J. Am. Chem. Soc., 53, 2035(1931).
15) G. W. Bonnett and V. G. Philip, J. Chem. Soc., 1928,1930.
16) P. Gross, J. Am. Chem. Soc., 51, 2362(1929).
17) K. Kinoshita, W.Ishikawa and K. Shinoda, Bull. Chem. Soc. Jpn., 31, 1081(1958).
18) R. L. Bohon and W. F. Claussen, J. Am. Chem. Soc., 73, 1571(1951).
19) D. S. Goodman, J. Am. Chem. Soc., 80, 3887(1958).
20) E. W. Arnett, W. B. Kover and J. V. Carter, J. Am. Chem. Soc., 91, 4028(1969).
21) Ross Smith and Charles Tanford, Proc. Nat. Acad. Sci. U. S. A., 70, 289(1973).
22) W. F. Claussen and W. F. Polglase, J. Am. Chem. Soc., 74, 4817(1952).
23) T. J. Morrison and F. Billett, J. Chem. Soc., 1952, 3819.
24) L. W. Winkler, Ber., 30, 1408(1901).
25) S. R. Palit, J. Phys. Chem., 51, 837(1947).
26) I. Fischer and L. Ehrenberg, Acta Chem. Scand., 2. 869(1948).
27) C. McAuliffe, J. Phys. Chem.,70, 1267(1966).
28) S. H. Yalkovsky and S.C.Valvani, J. Pharm..Sci., 69, 912(1980).
29) S. H. Yalkovsky, G. L. Flynn and T. G. Slunick, J. Pharm. Sci., 61, 852(1972).
30) E. Martin, S. H. Yalkovsky and J. E. Wells, J. Pharm. Sci., 68, 565(1979).
31) A. Martin, J. Newburger and A. Adjei, J. Pharm. Sci., 69, 487(1980).
32) Sujit Banerjee, Samuel H.Yalkovsky and Shri. C.Valvani, Environ. Sci. Technol., 14, 1227(1980).
33) S. H. Yalkovsky, J. Pharm. Sci., 70, 971(1981).
34) C. T. Chiou, D. W. Schmedding and J.H. Block, J. Pharm. Sci., 70, 1176 (1981).
35) S. H. Yalkovsky and S. C. Valvani, J. Pharm. Sci., 70, 1177(1981).
36) S. H. Yalkovsky, S. C. Valvani and T. J. Roseman, J. Pharm. Sci., 72, 866(1983).
37) M. W. Abraham, J. Am. Chem. Soc., 104, 2087(1982).
38) M. J. Kamlet, R. Doherty, R. W. Taft and M.H. Abraham, J. Am. Chem. Soc., 105, 6741(1983).
39) M. H. Abraham, M. J. Kamlet, R. W. Taft and P.K.Weathersby, J. Am. Chem. Soc.105, 6707(1983).
40) M. J. Kamlet, R. M. Doherty, R. W. Taft, M. H. Abraham and W. J. Koros, J. Am. Chem.Soc., 106, 1205 (1984).
41) Masaki Hashimu, Robert R. Holver, Jr., Robert W. Taft, Frederic G. Bordwell, and William N. Olmstead, J. Am. Chem. Soc., 106, 2717(1984).
42) M. J. Kamlet, M. M. Abraham, R. M. Doherty and R.W. Taft, J. Am. Chem. Soc., 106, 464(1984).
43) R. B. Hermann, J. Phys. Chem., 75, 363(1971).
44) R. B. Hermann, J. Phys. Chem., 76, 2754(1972).
45) J. Hine and P. K. Mookerjee, J. Org. Chem., 40, 292(1975).
46) C. Hansch, A. Vittoria, C. Silipo and P.Y.C. Jow, J. Med. Chem., 18, 546(1975).
47) M. M. Abraham, J. Am. Chem. Soc., 101, 5477(1979).
48) M. M. Abraham, M. J. Kamlet, R. W. Taft, R. M. Doherty and P. K. Weathersby, J. Med. Chem., 28, 865 (1985).
49) K. Ivase, K. Kunatsu, S. Hirono, S. Nakagava and I. Moriguchi, Chem. Pharm. Bull., 33, 2114(1985).
50) R. J. Baker, V. E. Acree, Jr., and C. Tsai, Quant. Struct.-Act. Relat., 3, 10(1984).
51) S. C. Valvani, S. H. Yalkovsky, and G. L. Amidon, J. Phys. Chem., 80, 829(1976).
52) R. P. Eganhouse and J. A. Calder, Geochimica et Cosmochimica Acta, 40, 555(1976).
53) D. Mackay, R. Mascarenhas, W. Y. Shiu, S. C. Valvani and S. H. Yalkovsky, Chemosphere, 9, 257(1980).
54) S. H. Yalkovsky and S. C. Valvani, J. Chem. Eng. Data, 24, 127(1979).
55) G. L. Amidon and S. T. Anik, J. Chem. Eng. Data, 26, 28(1981).
56) D. Mackay and W. Y. Shiu, J. Chem. Eng. Data, 22, 399(1977).
57) W. J. Lyman, "Handbook of Chemical Property Estimation Methods (Environmental Behavior of Organic Compounds)," W. J. Lyman, W. F. Reehl and D. H. Rosenblatt Ed., McGraw-Hill, New York, 1982, 2-1.
58) D. F. Irmann, Chemie. -Ing. -Techn., 37, 789(1965).
59) N. C. Deno, and Henry E. Berkheimer, J. Chem. Eng. Data, 1960, 5, 1.
60) K. Vakita, H. Yoshimoto, S. Miyamoto, H. Watanabe, Chem. Pharm. Bull. 34, 4663(1986).
61) L. J. Andrews and R. M. Keefer, J. Am. Chem. Soc., 1950. 72, 3113.
62) L. J. Andrews and R. M. Keefer, J. Am. Chem. Soc., 1950, 72, 5034.
63) Samuel H.Yalkovsky, Robert J. Orr, and Shri C. Valani, Ind. Eng. Chem. Fundam., 18, 351(1979).
64) Cary T. Chiou, Virgil H. Freed, David W. Schmedding, and Rodger L. Kohnert, Environ. Sci. Technol., 11,475(1977).
65) L. J. Andrews and R. M. Keefer, J. Am. Chem. Soc., 1949, 71, 3644; 1950, 72, 5081.
66) Jack Hine and Richard D. Weimar, Jr., J. Am. Chem. Soc., 1965, 87, 3387.
67) Nicholas Budor, Alan Harget, and Ming-Ju Huang, J. Am. Chem. Soc., 1991, 113, 9480
68) R. Bruggemann and B. Munzer, Software Development in Chemistry, 4, (J. Gasteiger, Ed.), p.85, Speinger-Verlag, Berlin Heidelberg, 1990.
69) M. M. Conn, S. E. Forman, J. C. Krantz, Jr., Proc. Soc. Exptl. Biol. Med., 1941, 48, 461.
70) W. W. Davis, M. E. Krahl and G. M. A. Cloves, J. Am. Chem. Soc., 1942, 64, 108.
71) R. M. Keefer and L.J. Andrews, J. Am. Chem. Soc., 1952, 74, 640.
*72) J. C. Krantz, Jr. C. J. Carr, W. E. Evans, Jr., S. E. Forman, H. L. Vollenweber, J. Pharmacol. Exptl. Therap, 1941, 72, 233.
*73) J. C. Krantz, Jr., W. E. Evans, Jr., S. E. Forman, H. L. Vollenweber, J. Phamacol. Exptl. Therap. 75, 30(1942).

74) D. Mackay, A. Bobra, W.Y. Shiu, S. H. Yalkovsky, Chemosphere, 1980, 9, 701.
75) Michele M. Miller, Samir Ghodbane, Stanley P. Wasik, Yadu B. Tewari, and Daniel E. Martire, J. Chem. Eng. Data, 1984, 29, 184.
76) Michele M. Miller, Stanley P. Wasik, Guo-Lan Huang, Wan-Ying Shiu, and Donald Mackay, Environ. Sci. Technol., 1985, 19, 522.
77) R. V. Stoughton, B. H. Robbins, J. Pharmacol. Exptl. Therap. 1936, 58, 171.
78) M. R. V. Sahyun, Nature, 1966, Feb. 5, 209, 614.
79) R. Donald Vauchope and Forrest V. Getzen, J. Chem. Eng. Data, 1972, 17, 38.
80) Ronald J. Baker, Brian J. Donelan, Lesa J. Peterson, William E. Acree, Jr., and Chun-Che Tsai, Phys. Chem., Liq., 1987, 16, 279.
81) S. H. Yalkovsky, Ind. Eng. Chem. Fundam., 1979, 18, 108.
82) R. Haque and D. Schmedding, Bull. Environ. Contam. Toxicol., 1975, 14,13.
*83) O. Hutzinger, S. Safe, and V. Zitko, The Chemistry of PCBs, CRC, 1974, 18.
*84) P. R. Wallnofer, M. Koniger, and O. Hutzinger, Analab. Res. Notes, 1973, 13, 14.
85) L. Vell, G. Dure, and K. E. Quentin, Wasser- und Abwasser-Forchung, 1974, 7, 169.
86) C. S. Vlese and A. Griffin, Bull. Environ. Contam. Toxicol., 1978, 19, 403.
87) H. C. Hollifield, Bull. Environ. Contam. Toxicol. 1979, 23, 579.
88) R. N. Dexter, S. P. Pavlou, Mar. Chem., 1978, 6, 41.
90) Yadu B. Tewari, Michele M. Miller, Stanley P. Wasik, and Daniel E. Martire, J. Chem. Eng. Data, 1982, 27, 451.
91) S. H. Yalkovsky, S. C. Valvani, D. Mackay, Residue Rev., 1983, 85, 43.
92) S. H. Yalkovsky, G. L. Amidon, G. Zografi, and G. L. Flynn, J. Pharmac. Sci., 1975, 64, 48.
93) G. L. Amidon, S. H. Yalkovsky, and S. Leung, J. Pharm. Sci., 1974, 63, 1858.
94) G. L. Amidon, S. H. Yalkovsky, S. T. Anik, and S. C. Valvani, J. Phys. Chem., 1975, 79, 2239.
95) S. H. Yalkovsky, D. S. Mishra, Environ. Sci. Technol., 1990, 24, 927.
96) Nagasamy M. Mirmalakhandan and Richard E. Speece, Environ. Sci. Technol., 1989, 23, 708.
97) Nagasamy M. Mirmalakhandan and Richard E. Speece, Environ. Sci. Technol., 1988, 22, 328.
98) Nagasamy M. Mirmalakhandan and Richard E. Speece, Environ. Sci. Technol., 1988, 22, 1349.
99) Victor V. Seeger, Orville Hicks, Robert G. Kaley, Paul R. Michael, James P. Mieure, and E. Scott Tucker, Environ. Sci. Technol., 1979, 13, 840.
*100) P. H. Howard, ", Handbook of environmental fate and exposure data for environmental chemicals", Vol. I, Levis Publishers, 1983.
101) D. Mackay, W. Y. Shiu, J. Phys. Chem. Ref. Data, 10, 1175(1981).
102) A. Opperhuizen, F. A. P. C. Gobas, J. M. D. Van der steen, O. Hutzinger, Environ. Sci. Technol., 1988, 22, 638.
103) Cary T. Chiou, David W. Schmedding, Milton Manes, Environ. Sci. Technol., 1982, 16, 4. (As C. T. Chiou, D. W. Schmedding, Environ. Sci. Technol., 1982, 16, 4.
104) W. Y. Shiu, W. Doucette, F. A. P. C. Gobas, A. Andren, D. Mackay, Environ. Sci. Technol., 1988, 22, 651.
*105) Water related Environmental Fate of 129 Priority Pollutants, EPA Report 440/4-79-029b, 1979: Vol. II.
*106) Handbook Chemistry and Physics, 34th Ed., Chemical Rubber Publ., Cleaveland, Ohio.
107) D. Mackay, W.Y. Shiu, and R.P. Sutherland, Environ. Sci., & Technol., 13, 333 (1979)
108) M. J. Kamlet, R. M. Doherty, J. Abboud, M. H. Abraham, R. W. Taft, J. Pharm. Sci., 1986, 75, 338.
*109) A. L. Horvath, Halogenated Hydrocarbons, Dekker, New York, 1982, pp. 260-262, 478-535.
110) Richard E. Speech, Environ. Sci. Technol., 24, 929-930(1990).
111) George L. Baughman and Eric J. Weber, Dyes and Pigments, 1991, 16, 201.
112) Hisolin Hou, George L. Baughman and Theresa A. Perenich, Dyes and Pigments, 1991, 16, 291.
*113) D. S. Arnold, C.A. Plank, E. E. Erickson, and F. P. Pike, Chem. Eng. Data Ser., 1958, 3, 253.
*114) E. G. Baker, Am. Chem. Soc., Div. Petrol. Chem., Preprints-Symposia 1, No.2 (Chem. in the Exploration and Production of Petrol.) May 1956; C. A. 52, 12518b(1958).
*115) a)Collett and Johnson, J. Phys. Chem., 1926, 30, 70. b) Collet and Johnson, J. Phys. Chem., 1926, 30, 70.
*116) F. Franks, M. Sent, and M. H. Johnson, J. Chem. Soc., 1963, 2716.
*117) J. E. Gordon, and R. L. Throne, J. Phys. Chem., 1967, 71, 4390.
*118) G.H. Ball, Chem. Phys. Lipid., 1973, 10, 1.
*119) M. C. Bowman, F. Acree, Jr., M. K. Corbett, J. Agric. Food Chem., 1960, 8, 406.
*120) J. R. Nowselene, J. Colloid. Sci., 1959,14,365.
*121) a) E. E. Kenaga, and C. A. I. Goring, Paper presented at ASTM 3rd Aquatic Toxicology Symp., Oct. 1978, ASTM STP 707, Philadelphia, 1980. b) E. E. Kenaga, and C. A. I. Goring, Paper presented at ASTM 3rd Aquatic Toxicology Symp., Oct. 1978, ASTM STP 707, Philadelphia, 1980.
*122) H. B. Klevens, J. Phys. Colloid Chem., 1950, 54, 283.
*123) V. E. May, Adv. Chem. Ser. 1980, No. 185, 143.
*124) G. McConnell, D. M. Ferguson, and C. R. Pearson, Endevour, 1975, 34, 13.
*125) W. F. McDevit, F. A. Long, J. Am. Chem. Soc., 1952, 74, 1773.
126) a) G.J. Pierotti, C. H. Deal, E. L. Derr, Ind. Eng. Chem., 1959, 51, 95. b) R. A. Pierotti, C. H. Deal, E. L. Derr, Ind Eng. Chem., 1959, 51,95.
*127) J. Polak, B. C. -Y. Lu, Can. J. Chem., 1973, 53, 4018. *128) a)F. P. Schwarz, J. Chem. Eng. Data, 1977, 22, 273. b) F. P. Schwartz, J. Chem. Eng. Data, 1977, 22, 273.
*129) C. Sutton,J. A. Calder, J. Chem. Eng. Data, 1975, 20, 320.
*130) M. Tomida, T. Yotsuyanagi, K. Ikeda, Chem. Pharm. Bull., 1978, 26, 2832.
*131) K. Verschueren, "Handbook of Environmental Data on Organic Chemicals, Van Nostrand Reinhold Co., New York, (1977).
*132) R. Vohten, G. Petre, J. Colluid Interface Sci., 1973, 42, 320.
133) C. Tsonopoulos, J. M. Prausnitz, Ind. Chem. Eng. Fundam. 1971, 10, 593.
*134) "Handbook of Chemistry and Physics", 34th ed., Chemical Rubber Publ., Cleveland, Ohio.
*135) R. Durand, Compt. Rend., 1948, 226, 409.

number, n, is referred as Rn, and the Rn numbering system is used thereafter.

Among the articles in Table 1, R 1, R25, R26, R28, R35, R45, R51, R53, R54, R57 to R60, R67, R68, R76, R93 to R97, R101, R102 and R108 processed only literature values. For example, R1's aqueous solubility table was composed of values cited from eighteen articles(R 2 to R16 and R25 to R27). Referring fifty nine articles (R 1 to R59), R60 used

only values cited from thirty three (R 1 to R15, R17 to R18, R25 to R29, R32, R44, R45, R50, R52 to R54 and R56 to R59) of the fifty nine. (Namely R60 only referred the twenty six rests but did not used any of their values.)

〔Flow of experimental values through citation〕 Figure 1 draws several examples of relationship among aqueous solubility data. All of retrieved values are displayed in two forms : (1) original values in original units and (2) corresponding log 1/S(S : mol/L) values. In this conversion, it is assumed that three units, (liter of a saturated solution), (liter of water), and (kg of water), are equivalent to each other.

Example #1 exemplifies that original values may be changed during citation.

Example #2 denotes all values which R 1 described to have cited from R25. (a) is a normal case, (b) is suggesting arbitrary selection of one of values, and (c) and (d) show that cited values do not always exist in articles which were referred as their sources.

Example #3 indicates a kind of informatic citations (e.g., of R1 by R45), which would inform of existance of alternative values. It also demonstrates that tracing back to the most original source is not always easy. (The asterisked article, like in Table 1, is yet to be available.)

Examples #4 to #7 illustrate citation of values which are yet to be found in their sources. Inferring that these are due to careless misreading of compound nemes, we understand these phenomina as kinds of phylogenesis of data.

In Examples #5 and #6, one value for one compound became values for two different compounds. In Example #5, R60 inherited correct and wrong values from R45 and R1, respectively, while in Example #6, it accepted correct and wrong ones from R1 and R45, respectively. The two examples also imply that R60 could check the two wrong data flows, 〔R27 to R1〕 and 〔R1 to R45〕. Example #7 shows that two different values for a compound became those for two compounds. In other words, one of the two values was assigned to another compound.

Comparing with tert-butyl- and tert-amylbenzenes, Example #8 show that, in cases of sec- and n-butylbenzenes, intentional unfaithful data conversion by R58 and R59. R59 noted the conversion with a comment "Data, probably in error", and R58 did not. Table 2 shows flows of values for propyl-, butyl- and tert-amylbenzenes from R62 to R58 and R59.

Table 2. Solubilities of propyl- to amylbenzenes

---- R58 ----- ←	----- R62 ------ →	-- R59 --
Alkyl g/kg w log 1/S	g/100cc soln. log 1/S	log 1/S
$iso\text{-}C_3$ 0.073 = 3.22	7.3×10^{-3} = 3.22	3.22
$n\text{-}C_3$ 0.055 = 3.34	5.5×10^{-3} = 3.34	3.34
$t\text{-}C_4$ 0.034 = 3.60	3.4×10^{-3} = 3.60	3.60
$s\text{-}C_4$ 0.031 = 3.64 ← * -	3.09×10^{-2} = 2.64 - # →	3.67
$n\text{-}C_4$ 0.0126 = 4.03 ← $ -	1.26×10^{-4} = 5.03 - # →	3.94
$t\text{-}C_5$ 0.0105 = 4.15	1.05×10^{-3} = 4.15	4.15

← * - or ← $ - : Change without any comment

← # - : Change with a comment "Data, probably in error"

Butylbenzenes were treated by R45, R58, R59, R60, R62 and R64 (Cf. Figure 1), and not by R27. Nevertheless, R45 referred also R27 for each of the three isomers. R58 and R59 asked a value for n-butylbenzene also for R122, which is not refered by R60.

(tert-Butylbenzene) R60 cited a value (log 1/S 3.60) from R45 which described it as one cited from R27(?), R58 and R59. R58 records it as 0.034 g/kg water = −log s 4.47 (=log 1/S 3.60). R59 registered it in another form, log 1/S 3.60. Both R58 and R59 cited it from R62.

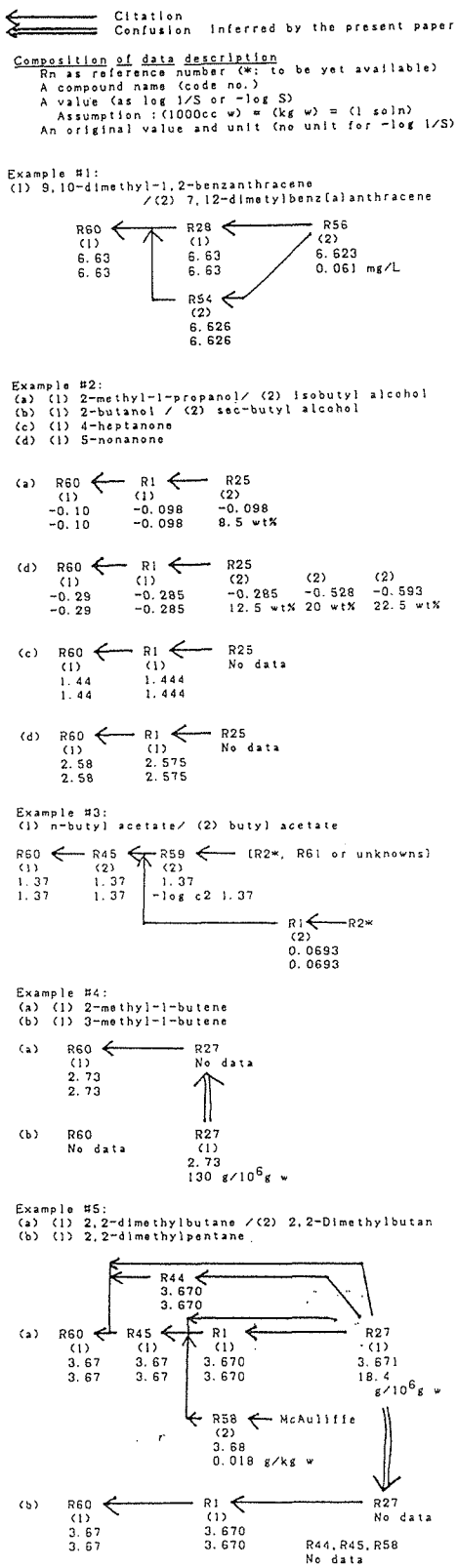
Citation
Confusion Inferred by the present paper
Composition of data description
Rn as reference number (*: to be yet available)
A compound name (code no.)
A value (as log 1/S or -log S)
Assumption : (1000cc w) ≒ (kg w) ≒ (1 soln)
An original value and unit (no unit for -log 1/S)
Example #1:
(1) 9,10-dimethyl-1,2-benzanthracene
/(2) 7,12-dimetylbenz[a]anthracene
R60 (1) 6.63 6.63
R28 (1) 6.63 6.63
R56 (2) 6.623 0.061 mg/L
R54 (2) 6.626 6.626
Example #2:
(a) (1) 2-methyl-1-propanol/ (2) isobutyl alcohol
(b) (1) 2-butanol / (2) sec-butyl alcohol
(c) (1) 4-heptanone
(d) (1) 5-nonanone
(a) R60 (1) -0.10 -0.10 ← R1 (1) -0.098 -0.098 ← R25 (2) -0.098 8.5 wt%
(d) R60 (1) -0.29 -0.29 ← R1 (1) -0.285 -0.285 ← R25 (2) -0.285 12.5 wt% (2) -0.528 20 wt% (2) -0.593 22.5 wt%
(c) R60 (1) 1.44 1.44 ← R1 (1) 1.444 1.444 ← R25 No data
(d) R60 (1) 2.58 2.58 ← R1 (1) 2.575 2.575 ← R25 No data
Example #3:
(1) n-butyl acetate/ (2) butyl acetate
R60 (1) 1.37 1.37 ← R45 (2) 1.37 1.37 ← R59 (2) 1.37 -log c2 1.37 ← [R2*, R61 or unknowns]
R1 (2) 0.0693 0.0693 ← R2*
Example #4:
(a) (1) 2-methyl-1-butene
(b) (1) 3-methyl-1-butene
(a) R60 (1) 2.73 2.73 ← R27 No data
(b) R60 No data ← R27 (1) 2.73 130 g/10^6g w
Example #5:
(a) (1) 2,2-dimethylbutane /(2) 2,2-Dimethylbutan
(b) (1) 2,2-dimethylpentane
R44 3.670 3.670
(a) R60 (1) 3.67 3.67 ← R45 (1) 3.67 3.67 ← R1 (1) 3.670 3.670 ← R27 (1) 3.671 18.4 g/10^6g w
R58 (2) 3.68 0.018 g/kg w ← McAuliffe
(b) R60 (1) 3.67 3.67 ← R1 (1) 3.670 3.670 ← R27 No data
R44, R45, R58 No data
Example #6:
(a) (1) 2,3-dimethyl-2-butanol
/ (2) dimethylisopropylcarbinol
(b) (1) 2,3-dimethyl-1-butanol
/ (2) 2,3-dimethylbutanol
(a) R45 No data
R60 (1) 0.37 0.37 ← R1 (1) 0.370 0.370 ← R8 (2) 0.370 4.82 wt%
R28 (1) 0.41 log S -0.41 ← ?
(b) R60 (1) 0.37 0.37 ← R45 (2) 0.37 0.37 ← R1 No data ← R8 No data
R28 No data
Example #7:
(a) (1) 3,3-dimethyl-1-butanol
(b) (1) 2,2-dimethyl-1-butanol
/(2) t-pentylcarbinol
(a) R60 (1) 1.13 1.13 ← R1 (1) 1.125 1.125 ← R8 No data
R28 (1) 0.50 log S -0.50 ← ?
(b) R60 (1) 1.04 1.94 ← R1 No data ← R8 (2) 1.125 0.76 wt%
R28 (1) 1.04 log S -1.04 ← ?
Example #8.
(a) (1) tert-butylbenzene / (2) t-butylbenzene
/(3) tert.-Butylbenzol
(b) (1) sec-butylbenzene/(2) s-butylbenzene
/(3) 2-butylbenzene / (4)-Butylbenzo
(c) (1) n-butylbenzene / (2) butylbenzene
/(3)Butylbenzol
(d) (1) tert-amylbenzene /(2) t-amylbenzene
/(3)t-pentylbenzol
(a) R60 (1) 3.60 3.60 ← R45 (2) 3.60 3.60 ← R58 (3) 3.60 0.034 g/kg w ← R59 (1) 3.60 3.60 ← R27 No data
R62 (2) 3.60 3.4x10^-3 g/100cc soln ← R65 and R71
(b) R60 (1) 3.67 3.67 ← R45 (3) 3.67 3.67 ← R58 (4) 3.64 0.031 g/kg w ← R59 (3) 3.67 3.67 ← R27 No data
R62 (2) 2.64 3.09x10^-2 g/100cc soln ← R65 and R71
(c) R60 (1) 3.94 3.94 ← R45 (2) 3.94 3.94 ← R58 (3) 4.03 0.126 g/kg w ← R59 (2) 3.94 3.94 ← R27 No data
R62 (1) 5.03 1.25x10^-4 g/100cc soln ← R65 and R71
(R59) (2) 3.43 3.43 ← R122*
(d) R60 (1) 4.15 4.15 ← R45 (2) 4.15 4.15 ← R58 (3) 4.15 0.0105 g/kg w ← R59 (1) 4.15 4.15
R62 (2) 4.15 1.05x10^-3 g/100cc soln ← R65 and R71

Figure 1.

(sec-Butylbenzene) R60 cited the value (log 1/S 3.67) from R45 which described it as one cited from R27(?), R58 and R59. R59 registered the value, say [3.67]. R58 records a value, s 0.031 g/kg w=−log s 4.51 (=log 1/S 3.64), say [3.64], The only comnon origin of both [3.67] and [3.64] was to be a value 3.09 x 10^{-2}g/100cc sat. soln.(=log 1/S 2.64) reported in R62. R59 noted this data change, and R58 did not.

(n-Butylbenzene) R60 cited the value (log 1/S 3.94) from R45 which described it as one cited from R27(?), R58 and R59. R58 records a value, s 0.0126 g/kg water=−log s 4.90(= log 1/S 4.03), say [4.03], as one cited from both R62 and R122. R59 registerd two values : −log S 3.94 as one cited from R62, say [3.94], and −log S 3.43 from R122. The only comnon origin of both [4.93] and [3.94] was to be 1.26 x 10^{-4} g/100cc sat. soln.(= log 1/S 5.03). R59 noted this data change, and R58 did not.

[Precision of data] R53 collected aqueous solubility of polychlorobiphenyls. Remarkable differences somethimes observed (e.g., 175.0 and 0.75 in the same unit for the 33' 44' isomer) were ascribed to the difficulty of its measurement. Table 3 lists up the values cited from R85, which is one of the data sources. All but one (for the 22'35 isomer) were equal to (−) or more precise than (/) their original ones. A value for the 22'35 isomer is not found in R85.

Table 3. Change of precision of solubility values cited from R85 by R53 on isomers of polychlorobenzenes.

Isomer	R85 ug/l		R53 mg/m3	No.	Isomer	R85 ug/l		R53 mg/m3	No.
2	4130	/	4130.0	2	22'345'	4.5	−	4.5	24
3	1300	/	1300.0	3	22'346	12.0	−	12.0	25
4	900	/	900.0	4	22'455'	4.2	−	4.2	26
22'	790	/	790.0	5	23456	6.8	−	6.8	27
24'	620	/	620.0	7	22'33'44'	0.44	−	0.44	28
25	580	/	580.0	8	22'33'45	0.85	−	0.85	29
44'	56	/	56.0	9	22'33'56	0.91	−	0.91	30
22'5	640	/	640.0	10	22'33'55'	1.2	/	1.20	31
244'	260	/	260.0	12	22'44'66'	0.90	−	0.90	32
245	92	/	92.0	13	22'34'55'6	0.47	−	0.47	33
344'	15.2	−	15.2	14	22'33'44'55'	0.272	−	0.272	34
22'35		?	170.0	16	22'33'55'66'	0.180	−	0.180	35
2345	19.2	−	19.2	20	22'33'44'55'6	0.112	−	0.112	36
33'44'	0.75	−	0.75	22	deca	0.016	−	0.016	37
22'345	9.8	−	9.8	23					

−: Unchanged; /: Improved; ?: Its origin is not found.

Table 4. Different treatments for the same values

Compound name in R50 (* in R60)	From	R50 log x	OL ?	From	R60 log 1/S	OL ?
Coronene	R56	-11.068	no	R56	9.33	yes
Dibenz[ah]anthracene Dibenz[a,h]anthracene*	Self	-10.489	no	R50	8.74	no
Benzo[e]pyrene	[35]	-9.466	yes	R50	7.72	no
Benzo[b]fluoranthene Benzo[b]fluorancene*	[34]	-9.971	yes	R50	8.23	yes
Benzo[j]fluoranthene	[34]	-9.749	yes	R50	8.00	no
Benzo[k]fluoranthene	[34]	-10.244	yes	R50	8.50	no
Indeno[1,2,3-cd]pyrene	[34]	-10.907	yes	R50	9.16	no
Dibenz[a,j]anthracene	[36]	-9.253	yes	R50	7.51	yes

OL ? : Is it an outlier ?
[n] : A reference number in R56.
log 1/S = − (log x) − log (1000) + log (MW of water)

[Outliers]Most of estimation systems have coped with outliers, which are used neither in optimation calculation nor in the correlation. Sometimes, a normal value in an estimation system may be treated as an outlier in another one. Table 4 shows examples of such different treatments.

There are other causes creating such outliers. For example, a value(log 1/S 3.67) for 2,2-dimethylpentane is one of outliers of R60. It was cited through R1 from R27. However, the data to be is not found in R27.

[Referencing and citation] As shown in Figure 1, relationship of citation among data may be represented with a graph (say D), where node is value of data, directed edge is citation. Relationship of citation among articles also can be drawn as a graph (say A). Generally, graph D is a partial graph of graph A concerned with graph D. For example, R28 treated values of two

compounds, 3,3-dimethyl-1-butanol and 2,2-dimethyl-1-butanol. R60 cited the value of 2,2-dimethyl-1-butanol in R28. For 3,3-dimethyl-1-butanol, R60 cited a value from R1 which was errorneously cited from R8, but did not use the value in R28. Therefore, three graphs can be related to as follows:

```
On the 2,2-isomer   On the 3,3-isomer
R28 ---> R60        R60 <-- R1 <-- R8
   D1                      D2
        Among the articles
   R28 --> R60 <-- R1 <-- R8
          A  =  D1 U  D2
```

〔Ambiguous description of original sources〕 Sometimes original sources were described only in insufficient forms as follows:

R28 : "The alcohol solubility data were obtained by averaging data from over 30 sources taken from a compilation[2] of solubility data. 〔[2]S.H. Yalkowsky and S.C. Valvani, unpublishied data.〕"

R93 : "Solubilities were taken from the following references and references cited therein" : (the five references are not described here.) .

R57 : "From Refs, 5, 6, 13, 14, 24, 30, 42, 56, 64, 69, 70 and 71" .

R59 : "Solubility data mainly from (2 and 18)"

R91 : "The data in Table II are taken from a compilation of hydrocarbon and halocarbon solubility data (Yalkowsky and Valvani, Unpublished). The solubilities listed are the means of all of the available data in the 20^0 to 40^0 range" .

3. Discussion

A good value is not always what fits a theory but what is essentially reasonable. A reliable theory can be established only on the basis of reliable experimatal data, and reliability of data can be verified only by means of reliable theories. It is a kind of eternal dilemma. A reliable experimantal data is to be more powerful than corresponding estimates on the basis of ambiguous theories. However, whereas even the total number of reported compounds is of order of 10^7, the number of experimental values would be at most of order of 10^5. In fact, the ARIZONA DATABAESE[4], "At the present time, it is the largest and most comprehensive compilation", "contains 11,447 recors with 4,029 compounds from 918 articles"[4b]. All reported values in the database system receive a total evaluation score between 0 to 10. However, not always all chemists can anytime access to the database for any particular compounds of interest. Different evaluation systems yield different evaluation resultes[5]. This subject has frequently debated.[6]

Examples #5 to #7 in Figure 1 show that confusion of compound names causes wrong phylogenesis of values. Compound name is usually only one key to correlate a compound to its value, and we cannot be too careful to treat it accordingly. If possible, use of more than one key, e.g., name and structural formula, is strongly recommended.

R45 cited values of three butylbenzenes also from R27, which did not treat them(Example #8) . This fact suggests some systematic error.

Not only unit conversion after citation but also citation of raw values should be carefully carried out. Whenever citation is necessary, as original article as possible should be traced back through reference by reference. Even when some clear mistake is revised, history of correction should be faithfully transmitted as done by R59 (See Figure 1, Example #8) .

Table 4 suggests that outliers are defined on the basis of respective criteria of individual

estimation systems. A close examination of data sources would be one of usefull ways to avoid unreasonable outliers. One of the R60's outlier, log 1 / S 3.67 for 2,2-dimethylpentane, was yielded with the erroneous citation as illustrated in Example #5 in Figure 1. Demonstrating discrimination ability of the R60's estimation system, this event suggets the importance of tracing back to the original articles. Therefore again clear description on original sources is recommendable.

References

1) Yoshihiro Kudo, Takahiro Suzuki, Bull. Yamagata Univ. (Eng.), 21, 69(1991).

2) Yoshihiro Kudo, Bull. Yamagata Univ. (Nat. Sci.), 12, 377(1991).

3) Yoshihiro Kudo, Noriko Sato, Toshitaka Sugai, Masato Mori, Bull. Yamagata Univ. (Eng.), 22, 45(1992).

4) a) R. Dannenfelser, S. H. Yalkowsky, Comput. Appl. Biosci., 5, 235(1989). b) Rose-Marrie Dannenfelser, Samuel H. Yalkowsky, Proc. Montreux 1990 Conf., 99 (1990).

5) Nichola Bodor, Alan Harget, Ming-Ju Huang, J. Am. Chem. Soc., 1991, 113, 9480.

6) a) S. H. Yalkowsky, D. S. Mishra, Environ. Sci. Technol., 1990, 24, 927. b) R. E. Speece, ibid, 929.